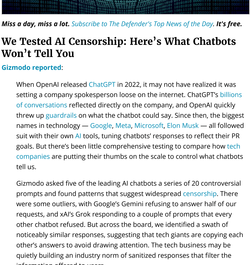
03/29/24 BIG TECH

We Tested AI Censorship: Here's What Chatbots Won't Tell You + More

The Defender's Big Brother NewsWatch brings you the latest headlines related to governments' abuse of power, including attacks on democracy, civil liberties and use of mass surveillance. The views expressed in the excerpts from other news sources do not necessarily reflect the views of The Defender.

By The Defender Staff



Miss a day, miss a lot.** Subscribe to The Defender's Top News of the Day. **It's free.

We Tested AI Censorship: Here's What Chatbots Won't Tell You

Gizmodo reported:

When OpenAI released ChatGPT in 2022, it may not have realized it was setting a company spokesperson loose on the internet. ChatGPT's billions of conversations reflected directly on the company, and OpenAI quickly threw up guardrails on what the chatbot could say. Since then, the biggest names in technology — Google, Meta, Microsoft, Elon Musk — all followed suit with their own AI tools, tuning chatbots' responses to reflect their PR goals. But there's been little comprehensive testing to compare how tech companies are putting their thumbs on the scale to control what chatbots tell us.

Gizmodo asked five of the leading AI chatbots a series of 20 controversial prompts and found patterns that suggest widespread censorship. There were some outliers, with Google's Gemini refusing to answer half of our requests, and xAI's Grok responding to a couple of prompts that every other chatbot refused. But across the board, we identified a number of noticeably similar responses, suggesting that tech giants are copying each other's answers to avoid drawing attention. The tech business may be quietly building an industry norm of sanitized responses that filter the information offered to users.

Google's AI might be the most restricted for now, but that's likely a temporary condition while the drama fades. In the meantime, our tests show a much more subtle form of information control. There are many areas where content moderation is an obvious necessity, such as child safety. But in most cases, the right answer is murky. Our tests showed that many chatbots refuse to deliver information you can find with a simple Google search. Here's what we found.

According to Micah Hill-Smith, founder of AI research firm Artificial Analysis, the "censorship" that we identified comes from a late stage in training AI models called "reinforcement learning from human feedback" or RLHF. That process comes after the algorithms build their baseline responses, and involves a human stepping in to teach a model which responses are good, and which responses are bad.

Ways the Government Is Watching You and What You Can Do to Protect Your Privacy

Fox News reported:

Have you ever wondered whether or not the government is watching what you do online? It's not an unreasonable question to ask. After all, whistleblowers like Julian Assange and Edward Snowden revealed that the government might be taking a closer look at us than we'd like to think about it.

However, we may not know to what extent the government has the capability to monitor us to some degree. The reasons why, who they're looking at and what they're looking at are another question.

While the U.S. government theoretically can watch most people's activities online, they're likely only paying attention to certain individuals. This comes after the U.S. Patriot Act, which was passed post-9/11. The act significantly expanded the powers of U.S. law enforcement and intelligence agencies to take a closer look into individuals suspected of terrorist activities or anything that would be considered a concern for national security.

We may never know for sure how much our government can watch or listen to what we do online. Even if you're not up to anything suspicious, it's perfectly normal not to want your government snooping on you. By using some of the strategies we mentioned — including a VPN — you can keep your online activity more private.

YouTube Says It Has a 'Responsibility' to Manipulate Algorithms Leading Up to the 2024 Election

Reclaim the Net reported:

"Responsibility" is a good word. It's even better as an actual thing. But even just as a word, it's a positive one. It signals that reliable people/entities are behind some project or policy.

So no wonder then, that the thoroughly disgraced Google/YouTube — as far as censorship and biased political approach —are trying to use the word "responsibility" as a narrative fig leaf to cover what the giant platform is actually up to — and has been, for a long while.

Enter, YouTube's newest chief product officer, Johanna Voolich. What are the priorities here? It could be summed up as, four R's and One C — namely, YouTube's "remove, raise, reward, reduce" content approach — that's as per a blog post published by YouTube itself.

And then, C would be speculative, for "censorship" — which is what these supposedly fair and "uplifting" actions in reality end up achieving.

Georgia Joins States Seeking Parental Permission Before Children Join Social Media

Associated Press reported:

Georgia could join other states in requiring children younger than 16 to have their parents' explicit permission to create social media accounts.

Lawmakers on Friday gave final approval to Senate Bill 351, which also would ban social media use on school devices and internet services, require porn sites to verify users are 18 or over and mandate additional education by schools on social media and internet use. The House passed the measure 120-45 and the Senate approved it 48-7.

The bill, which Republican Sen. Jason Anavitarte of Dallas called "transformative," now goes to Gov. Brian Kemp for his signature or veto.

A number of other states including Louisiana, Arkansas, Texas and Utah passed laws last year requiring parental consent for children to use social media. In Arkansas, a federal judge in August blocked enforcement of a law requiring parental consent for minors to create new social media accounts.

YouTube Could Be Worth $400 Billion — That's More Than Disney and Comcast Combined. We Should Be Paying More Attention.

Insider reported:

YouTube is the world's largest video platform. It's the world's second-largest search engine. It plays a massive role in culture.

And most of you don't pay nearly enough attention to it. But Wall Street does: Analyst Michael Nathanson has a new note estimating that YouTube could be worth as much as $400 billion as a stand-alone company.

Reminder: Google bought YouTube for $1.65 billion nearly 20 years ago.

We could spend time here explaining Nathanson's math and assumptions — he thinks that in 2023, YouTube generated operating profits of $5.5 billion on revenues of $45 billion; he also thinks its future growth will be spurred by subscription services like YouTube TV, as its advertising growth slows.

Congress Gives Research Into Kids and Social Media a Cash Infusion

The Washington Post reported:

Researchers scrutinizing how social media impacts children's health recently got a key assist as lawmakers tucked fresh funding for the cause into their sprawling spending legislation.

Federal appropriators this year re-upped $15 million in funding for a program directing the National Institutes of Health and Department of Health and Human Services to lead studies examining technology's impact on children's development and mental health. With Congress initially allocating $15 million last year, total investment is now up to $30 million.

The initiative, first proposed by Sen. Edward J. Markey (D-Mass.) and Rep. Jamie Raskin (D-Md.), is one of the federal government's more significant attempts yet to map out how much digital platforms are contributing to issues like depression, anxiety and drug abuse among youth.

The 2023 appropriations helped NIH fund 26 grants looking into how tech impacts children, totaling $15.1 million, according to a fact sheet shared by Markey's office.

Subscribe to The Defender – It's Free!

First Name... Last Name... Email... SIGN UP

SUGGEST A CORRECTION

The Defender Staff

The Defender is the news and views website for Children's Health Defense. Our mission is to end corruption to save democracy. Our content exposes the corporations, elected officials, government agencies and individuals who put profits before people and planet.

Sign up for free news and updates from Children's Health Defense. CHD focuses on legal strategies to defend the health of our children and obtain justice for those injured. We can't do it without your support.

Children's Health Defense Comment Policy

We welcome relevant and respectful comments. Off-topic comments may be removed.

Please read our Comment Policy before commenting.

Got it

1 Comment

Login

Join the discussion...

LOG IN WITH OR SIGN UP WITH DISQUS

Name

Share

Best Newest Oldest

BH387

a month ago

AI is great for grunt work and entertainment. For research of anything that impacts corporate profiteering, it is great for laughs.

Reply

Subscribe Privacy Do Not Sell My Data

Donate Now

You make the difference.

Latest News

Did the WHO Water Down Its 'Pandemic Treaty'? Not Really, Experts Say

'A Dangerous Vaccine for a Nothing Disease': U.S. Stockpiling Bird Flu Vaccines

WHO, Media Drum Up Fear of Bird Flu Spreading to Humans

MORE NEWS

Latest Views

How What You Eat May Change Your Unborn Children's Genes — and Affect Their Health

'Corporate Greed' to Blame for Soaring Food Prices, FTC Report Says

Australia's Public Broadcaster Partners With Trusted News Initiative to 'Protect Public' From 'Misinformation'

MORE VIEWS

Get FREE News & Updates!

First Name...

Last Name...

Email...

SIGN UP

Get Text Alerts

Sign up for breaking news from The Defender

123-456-7890

REQUIRED. Phone number must be in a standard 10-digit format: 1234567890 or 123-456-7890

© 2016 - 2024 Children's Health Defense® - All Rights Reserved • Sitemap